Theor Appl Genet (1997) 95: 842–849

F. A. Gaiotto · M. Bramucci · D. Grattapaglia

Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers

Received: 25 March 1997 / Accepted: 13 May 1997

Abstract *Eucalyptus* breeding is typically conducted by selection in open-pollinated progenies. As mating is controlled only on the female side of the cross, knowledge of outcrossing versus selfing rates is essential for maintaining adequate levels of genetic variability for continuous gains. Outcrossing rate in an open-pollinated breeding population of *Eucalyptus urophylla* was estimated by two PCR-based dominant marker technologies, RAPD and AFLP, using 11 open-pollinated progeny arrays of 24 individuals. Estimated outcrossing rates indicate predominant outcrossing and suggest maintenance of adequate genetic variability within families. The multilcous outcrossing rate (t_m) estimated from RAPD markers (0.93 ± 0.027), although in the same range, was higher ($\alpha > 0.01$) than the estimate based on AFLP (0.89 ± 0.033). Both estimates were of similar magnitude to those estimated for natural populations using isozymes. The estimated Wright's fixation index was lower than expected based on t_m possibly resulting from selection against selfed seedlings when sampling plants for the study. An empirical analysis suggests that 18 is the minimum number of dominant marker loci necessary to achieve robust estimates of t_m. This study demonstrates the usefulness of dominant markers, both RAPD and AFLP, for estimating the outcrossing rate in breeding and natural populations of forest trees. We anticipate an increasing use of such PCR-based technologies in mating-system studies, in view of their high throughput and universality of the reagents, particularly for species where isozyme systems have not yet been optimized.

Communicated by P. M. A. Tigerstedt

F. A. Gaiotto · D. Grattapaglia (✉)
Laboratório de Genética de Plantas – CENARGEN – EMBRAPA
C.P. 02372 Brasília, DF 70849-970 Brazil

F. A. Gaiotto · M. Bramucci
Depat. de Genética – UNESP – Campus de Botucatu
Botucatu, SP 18618-000 Brazil

Key words Mating system · AFLP · RAPD · *Eucalyptus*

Introduction

Long-term breeding is aimed at achieving a balance between continuous genetic gains and maintenance of adequate levels of genetic variation. The most common method in long-term forest-tree breeding is recurrent selection; i.e., choosing the best trees in the current population as parents of the next generation. This cycle is repeated every generation to increase the frequency of favorable alleles at loci that control economically important traits (Namkoong et al. 1988). Critical decisions in recurrent selection breeding include the choice of mating design and selection method, the intensity of selection applied, and proper management of inbreeding. At the same time, funding constraints for the program and the biological characteristics of the species have to be considered.

In *Eucalyptus*, the use of hybrids between some species in exotic conditions frequently results in significant genetic gains possibly due to either true heterotic effects and/or a combination of traits (Zobel et al. 1987). This feature has led an increasing number of forestry companies throughout the world, and particularly in Brazil, to adopt reciprocal recurrent selection as the long-term breeding strategy of choice. However, due to the difficulty in carrying out large numbers of controlled pollinations in *Eucalyptus*, the main populations of such a program are typically advanced through open-pollination and controlled crosses are used solely in the reciprocal phase. This is a simple and low-cost option, allowing rapid turnover of generations and providing for substantial genetic progress. However, an incomplete pedigree design has the inconvenience that male parentage is

assumed to be random and outcrossed. The breeder does not know whether the individuals selected were produced by self pollination, or are in fact half-sibs. Information on the predominant system of mating in such a breeding population is therefore essential as it ultimately regulates the distribution of genetic variation between and within families and determines the rate of increase in co-ancestry. Furthermore, it has been shown in *Eucalyptus*, that assuming open-pollinated families as true half-sib families might result in a severe bias when estimating genetic parameters and parental breeding values, particularly for material from native stands (Hodge et al. 1996).

Outcrossing rates in forest-tree populations have traditionally been estimated from polymorphism data at isozyme loci (Adams 1983). In recent years, however, several high-throughput PCR-based technologies have been developed to assay genetic polymorphism at the DNA level. Among these, RAPD (Williams et al. 1990) and more recently AFLP (Vos et al. 1995) have been increasingly used for detailed genetic analysis. Both technologies are accessible and quickly provide large numbers of polymorphic markers with universal reagents and assay protocols. In *Eucalyptus* genetics these markers have proven useful for several applications including linkage and QTL mapping (e.g., Grattapaglia and Sederoff 1994; Gaiotto et al. 1997). However, due to their dominant behavior, RAPD and AFLP markers provide less information per locus than co-dominant markers. This is particularly relevant for applications that require genotype discrimination, as in the case of outcrossing-rate estimation. Through simulation studies, Ritland and Jain (1981) demonstrated, however, that this limitation could be readily overcome by multilocus estimation using a large number of dominant markers with intermediate gene frequencies.

Natural populations of commercial species of *Eucalyptus* have a predominantly allogamous habit. The outcrossing rates for some *Eucalyptus* species in Australia were estimated to be between 0.69 and 0.86 (Moran and Bell 1983). For natural populations of *Eucalyptus urophylla*, the average outcrossing rate is 0.9 (House and Bell 1994). For populations in exotic conditions data are extremely scarce. Only one study to-date reported an outcrossing rate of 0.853 for a seed orchard of *Eucalyptus citriodora* in Brazil (Yeh et al. 1983). The present study had therefore the following objectives: (1) to test the utility of dominant markers, both RAPD or AFLP, for estimating outcrossing rates in forest-tree populations; (2) to empirically determine the adequate number of dominant marker loci necessary to achieve robust estimates of outcrossing rate; and (3) to estimate the relative frequencies of outcrossing versus selfing in a long-term breeding population of *E. urophylla* in an exotic condition in Brazil.

Material and methods

Plant material

Seeds of *E. urophylla* were originally collected in natural populations in Timor Island, and a provenance/progeny trial was established in Brazil by Aracruz Celulose S.A. One-hundred and twenty one trees were selected from this progeny trial to establish a long-term breeding population. Open-pollinated (O.P.) seeds were collected from these 121 selected trees. The maternal trees were then cut down and therefore were not available for genotype determination at the time of the present study. From the 121 O.P. families, a random sample of 11 O.P. progeny arrays of 24 individuals, i.e., a total of 264 individuals, were used for this study.

RAPD and AFLP marker analysis

Genomic DNA extraction and RAPD assay conditions were as described earlier for *Eucalyptus* (Grattapaglia and Sederoff 1994). Arbitrary primer screening was done with 24 ten-base primers (Operon Technologies) previously used for map construction (Grattapaglia and Sederoff 1994). Primers were screened using 16 random DNA samples of four individuals from each of four families to select primers that amplified common markers across families. Four primers were selected based on the number of RAPD polymorphisms amplified, their size, amplification intensity, and the presence of polymorphic markers across families. The RAPD fragment sizes were estimated by comparing the band mobility with fragment size standards. Each RAPD primer selected (Table 1) generated a profile of 10–15 bands from which an average of six were chosen to be used as markers. Thus, the four RAPD primers amplified a total of 23 dominant markers, which were employed for the estimation of outcrossing rate.

AFLP technology was used following the procedures described by Vos et al. (1995) employing *Eco*RI and *Mse*I as rare- and frequent-cutter enzymes, respectively. An optimization for species of *Eucalyptus* was done, mainly for the pre-amplification step, resulting in higher reproducibility in the final amplification step. The modification consisted in keeping the annealing temperature constant at 60°C throughout the PCR program. In the original program (Vos et al. 1995) the annealing temperature started at 65°C and decreased 0.7°C in the first cycle, until it was kept constant at 56°C in the last 23 cycles. Comparing the pre-amplification products on agarose gels, the modified cycle clearly generates a more consistent smear of bands than the original cycle. Final selective amplifications were tested with combinations of primers containing either two or three selective nucleotides at the 3′ end. Amplifications with primers having two selective nucleotides at the *Eco*RI end and

Table 1 Nucleotide sequences of selected RAPD and AFLP primers and the number of dominant markers amplified

Primer code	Sequence (5′ → 3′)	# dominant markers amplified
RAPD OPK-9	CCCTACCGAC	7
RAPD OPW-11	CTGATGCGTG	8
RAPD OPX-15	CAGACAAGCC	4
RAPD OPY-20	AGCCGTGGAA	4
AFLP E37	GACTGCGTACCAATTCACT	
AFLP M51	GATGAGTCCTGAGTAACCA	26

three selective nucleotides at the *MseI* end (+2/+3), or three selective nucleotides for both ends (+3/+3), were tested. The results showed that a +3/+3 primer combination gave less, but more robust, bands ensuring more precision when markers were scored. The +2/+3 primer combination gave a high background on the gel. The "E primers" are those that amplified fragments by annealing to the adapters ligated to the *Eco*RI sites; the "M primers" correspond to the *Mse*I sites. After screening 20 different primer combinations, one that generated 26 AFLP markers was chosen (Table 1). Both RAPD and AFLP markers were identified by the primer code followed by the estimated size of the fragment in base pairs, e.g., W11_1500 or E37M51_510.

Data analysis

RAPD and AFLP markers are typically dominant. Scoring of bands was done considering only two possible alleles: band presence or band absence. Single-locus (t_s) and multilocus (t_m) outcrossing rates, RAPD and AFLP marker-allele frequencies, and Wright's fixation index (F) were estimated according to the mixed-mating model (mixed outcrossing and selfing; Ritland and Jain 1981). This model specifies that both selfing and outcrossing occur in the population (Shaw and Allard 1982). Ritland's modification of the MLT program, know as MLDT (Multilocus estimation of t using dominant markers) (Ritland 1990), was used for the analysis. Variances of parameter estimates were found by numerical re-sampling employing the bootstrap method. A minimum of 100 bootstraps were used in the analysis. For all RAPD and AFLP marker loci employed in the estimation of outcrossing rates, maximum-likelihood estimates of allelic frequencies were obtained. Standard errors of allelic-frequency estimates were obtained by 100 bootstraps within families. For each locus, a χ^2 statistic was calculated to test the null hypothesis that the number of observed progeny individuals for each genotype class from each maternal genotype plant did not differ from the expected number under the mixed-mating model.

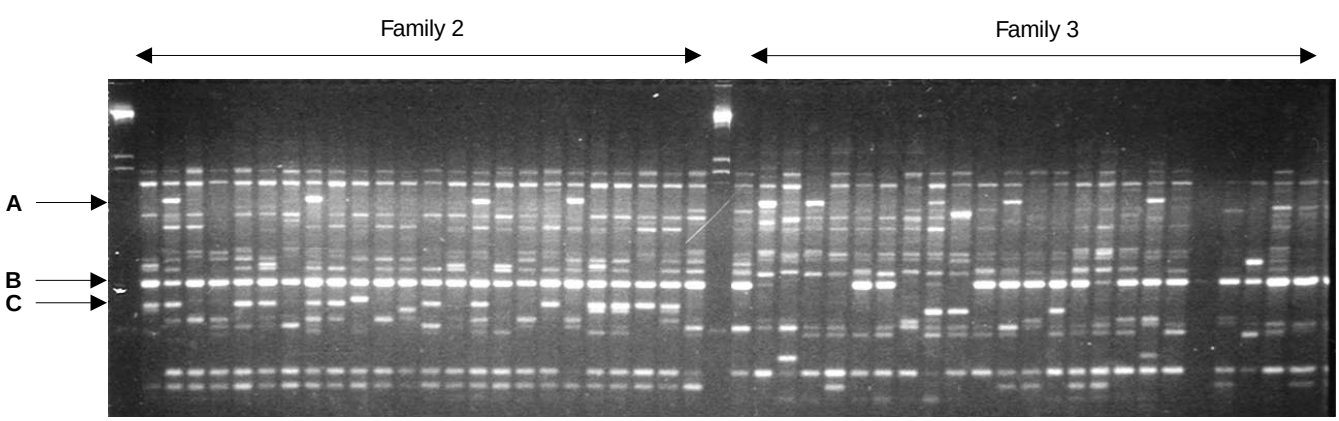
The estimates of t_m and t_s for AFLP and RAPD markers were compared and tested for significance using a paired *t* test. An empirical analysis was carried out to estimate the minimum number of dominant markers necessary to obtain a robust outcrossing-rate estimate. These analyses were done for each kind of marker separately. Beginning with four markers taken at random, the estimation of t_m was carried out by increasing by two the number of markers used in the analysis until the total available number of markers was used; 23 for RAPD and 26 for AFLP. The coefficient of variation on the estimate of t_m was computed and plotted with increasing number of markers.

Fig. 1 RAPD marker patterns in O.P. families. RAPD gels generated with primer OPW11 in family 2 (*lanes* 2–25) and 3 (*lanes* 27–50). Marker *A*: rare in both families; marker *B*: monomorphic in family 2 and frequent in family 3; marker *C*: frequent in, family 2 and rare in family 3. *Lanes* 1 *and* 25 are *Hin*dIII-digested DNA fragment size standards

Results

When using dominant AFLP or RAPD markers, the homozygous genotype for the dominant-allele "band presence" (+/+) and heterozygous genotype (+/−) cannot be distinguished from each other. In this case, the more informative markers for outcrossing-rate estimation are those where the dominant allele ("band presence") is at low frequency in the population, increasing the probability that the maternal plants are homozygous for "band absence", i.e. they do not have the RAPD or AFLP marker so that outcrossed offspring that have the dominant allele can be directly counted. Typically we observed that some dominant alleles can be frequent in an open-pollinated family and rare in another one (Figs. 1 and 2).

A RAPD or AFLP band present in all offspring (e.g., marker B in family 2, Fig. 1 and marker C in family 1, Fig. 2) indicates that the maternal plant could be homozygous for this marker (genotype +/+) or heterozygous or homozygous null, and that the marker could be at a very high frequency in the pollen pool, i.e., in the population. In all these cases the marker configuration is either of very little use, or else is non-informative, for detecting outcrossing events. Markers absent in a few offspring (e.g., marker B in families 2 and 3, Fig. 2) suggest that the maternal genotype could be either heterozygous (+/−) or homozygous null, with the marker at high frequency. In this case, offspring originating from outcrossing cannot be discriminated against based solely on this marker. Finally, the most-informative configuration is when a marker is present in just a few offspring (e.g., marker A in families 1,2 and 3 and marker B in family 1, Fig. 2). In this case it can be inferred that the maternal plant has a homozygous genotype (−/−) for the marker, and



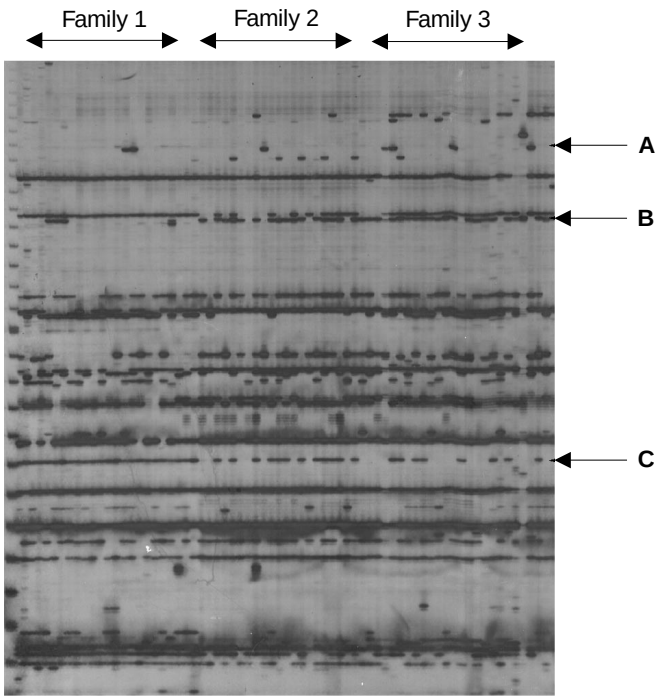


Fig. 2 AFLP marker patterns in O.P. families. Autoradiogram of AFLP markers (*A*, *B*, *C*) in three families with 24 individuals each. AFLP fragments were generated by $a + 3/ + 3$ primer combination (E37/M51) with "ACT" and "CCA" selective nucleotides respectively. The *first lane* on the left is a fragment size standard

that progeny individuals that have the marker are most surely heterozygous ($+/-$) and a product of outcrossing. So outcrossing events can be readily identified if the maternal plant has a homozygous null genotype. Evidently even in this case, not all of the outcrossing events can be identified based only on this marker, as progeny individuals that were outcrossed but received the "band-absent" allele from the paternal plant are not directly detectable. It is by accumulating data from several markers and estimating allele frequencies in the population for the RAPD or AFLP markers used, that the most-likely maternal genotypes can be inferred and all the outcrossing events detected.

Allele frequencies for the dominant "band-presence" allele and their standard errors were estimated by MLDT both for RAPD and AFLP assays. A χ^2 statistic to test the conformity of marker loci to the mixed-mating model, indicated that for six RAPD markers and five AFLP markers the number of observed progeny individuals for each genotype class from each maternal genotype departed from the expected numbers (Table 2). Because we did not have the maternal plants to assay, the potential maternal genotypes were inferred based on the dominant marker-allele frequency in the progeny array (Table 3).

The estimates of multilocus outcrossing rates (t_m) and single-locus outcrossing rates (t_s) obtained from MLDT, both from RAPD and AFLP data, clearly indicate that outcrossing is predominant in the *E. urophylla* open-pollinated breeding population. The multilocus outcrossing rate estimates based on all 23 RAPD loci and 26 AFLP loci were around 90%. Multilocus estimates did not significantly differ from the single-locus estimates. A Student's *t* statistics indicated significant difference between the estimates of t_m and t_s obtained from RAPD and AFLP data (Table 4).

An empirical analysis indicated that both for RAPD and AFLP, from 18 markers on, the estimates of t_m stabilized at around 0.9 and did not change significantly with an increasing number of markers. A smoother stabilization of the estimate of t_m was observed for RAPD data in comparison to AFLP data. The standard errors of the estimates, and hence the coefficient of variation (CV%), also rapidly decreased, as expected (Fig. 3). With less than ten markers, particularly for AFLP data, the estimates of t_m varied widely between 0.7 and 1.0. The coefficient of variation decreased from around 10–12% for 4–6 markers to around 4% with 18–26 markers. Finally, when 100 different random samples of 18 markers, either RAPD or AFLP, were taken to estimate t_m, the estimate always converged to a value close to 0.9 (data not shown).

Discussion

AFLP technology for *Eucalyptus*

This is the first report using AFLP markers for genetic analysis in *Eucalyptus*. Given that the genome size of commercial species of *Eucalyptus* does not vary significantly from species to species (Grattapaglia and Bradshaw 1994), the optimized protocol described in this study using three selective nucleotides at both priming ends should be readily transferable to other *Eucalyptus* species.

Our results indicate that the AFLP marker assay in *Eucalyptus* under optimized conditions is extremely reproducible and highly polymorphic, generating a large number of markers. Careful selection of primer combinations allows the generation of a large set of polymorphic markers in a single gel lane. Interpretation of data when markers across gels need to be scored, was found to be easier and more reliable with AFLP markers when compared to RAPD markers. Extra care and a higher stringency in selecting markers had to be taken when scoring RAPD markers across gels to avoid scoring errors with co-migrating or faint bands. The higher robustness of the AFLP assay compared with the RAPD assay was expected because specific PCR with longer primers is carried out at higher stringency. In the pre-amplification we used a higher and constant annealing temperature of 60°C instead of a touch-down protocol as described by Vos et al. (1995). This modification resulted in a higher

Table 2 Dominant marker loci, estimates of the dominant marker allele frequencies (+), their respective standard deviations (σ) and χ^2 statistics for agreement with the mixed-mating model

RAPD marker locus	Freq. (+)	σ	χ^2	AFLP marker locus	Freq. (+)	σ	χ^2
W11_1500	0.442	0.040	4.29*	E37M51_510	0.152	0.022	2.46
W11_1300	0.185	0.029	1.44	E37M51_450	0.061	0.013	1.37
W11_800	0.408	0.046	2.53	E37M51_385	0.078	0.021	2.10
W11_750	0.663	0.033	0.83	E37M51_365	0.020	0.009	1.34
W11_700	0.085	0.021	2.51	E37M51_320	0.073	0.017	2.67
W11_650	0.078	0.020	1.43	E37M51_317	0.025	0.009	0.57
W11_550	0.755	0.034	0.02	E37M51_297	0.113	0.021	2.06
W11_510	0.263	0.048	0.27	E37M51_295	0.016	0.007	2.89
K9_2600	0.690	0.032	0.62	E37M51_272	0.042	0.014	1.16
K9_1630	0.156	0.025	11.10**	E37M51_255	0.255	0.046	9.09**
K9_1050	0.151	0.033	0.17	E37M51_253	0.266	0.041	0.21
K9_1018	0.378	0.052	0.11	E37M51_215	0.465	0.044	1.32
K9_700	0.073	0.020	3.63	E37M51_207	0.222	0.031	10.56**
K9_480	0.751	0.037	1.30	E37M51_190	0.328	0.038	3.91*
K9_150	0.043	0.010	6.71**	E37M51_188	0.066	0.014	1.26
X15_1630	0.084	0.022	0.23	E37M51_182	0.104	0.020	0.58
X15_950	0.263	0.032	1.77	E37M51_173	0.051	0.021	5.83*
X15_750	0.073	0.020	0.71	E37M51_171	0.419	0.047	6.48*
X15_600	0.119	0.028	0.35	E37M51_160	0.110	0.021	4.04
Y20_1100	0.063	0.016	4.79*	E37M51_150	0.445	0.043	3.07
Y20_800	0.164	0.032	0.18	E37M51_125	0.300	0.036	2.41
Y20_550	0.350	0.044	4.15*	E37M51_124	0.089	0.020	2.21
Y20_400	0.277	0.041	7.69**	E37M51_117	0.020	0.009	0.25
				E37M51_99	0.096	0.023	0.17
				E37M51_90	0.385	0.040	0.24
				E37M51_89	0.064	0.016	1.00

* Marker locus with significant deviation at the 0.05 level
** Marker locus with significant deviation at the 0.01 level

reproducibility in the final amplification, possibly as a result of a much more rigorous selection of fragments at the pre-amplification stage, so decreasing the background in the detection of the final amplification products.

Estimation of outcrossing rate with dominant RAPD and AFLP markers

Dominant RAPD markers were previously used to estimate the outcrossing rate in populations of a flowering plant (Fritsch and Rieseberg 1992). The present study shows the use of dominant AFLP markers in a mating-system study and demonstrates that both RAPD and AFLP technologies, although supplying dominant markers with a lower information content than traditional co-dominant isozymes, are very adequate for the study of the mating system in populations and provide comparable results. These high-throughput marker technologies allow the analysis of a large number of individuals with a large number of markers in a relatively short time as only a few RAPD primers or a single AFLP primer pair allow the generation of sufficient markers to obtain a robust estimate of outcrossing rate. This study also points to the fact that when using dominant markers for outcrossing rate estimation it is imperative that an adequate screening of primers be done so as to maximize the probability of amplifying a large number of polymorphic markers in the progeny arrays. Only by amplifying a large number of markers does one have the flexibility of selecting highly reproducible ones that can be readily scored across families and which display low-to-intermediate frequency for the dominant marker allele in the population. The predominance of heterozygous (2) and homozygous null (3) classes for the inferred maternal genotypes (Table 2) reflects the selection of low-to-intermediate frequency markers necessary to carry out such a study.

The estimates of outcrossing rates obtained indicate that the open-pollinated breeding population of *E. urophylla* is preferentially outcrossing (Table 4) suggesting that, in fact, expected levels of genetic variation for open-pollinated families are likely to be maintained. Multilocus estimates did not differ significantly from the single-locus estimates, suggesting no significant biparental inbreeding. Both multilocus and single-locus outcrossing rates estimated both from RAPD and AFLP data were around 90%. These estimates are similar to that obtained using three isozyme loci in a seed orchard of *E. citriodora* in an exotic condition in Brazil, where a t_m of 0.853 was found (Yeh et al. 1983), and also closely agree with estimates for natural

Table 3 Maternal genotypes for the 11 O.P. families inferred by MLDT for each dominant marker. Genotype 1: homozygous for "band-presence" allele (+ / +); genotype 2: heterozygous (+ /); genotype 3: homozygous (− / −)

RAPD marker locus	O.P. Family											AFLP marker locus	O.P. Family										
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
W11_1500	2	3	3	3	2	3	1	3	2	1	2	E37M51_510	3	3	3	3	3	3	2	1	3	1	3
W11_1300	3	3	3	2	2	3	3	3	3	3	3	E37M51_450	3	3	3	3	3	3	3	3	3	3	3
W11_800	2	3	2	2	2	2	2	3	2	2	2	E37M51_385	3	3	3	3	3	3	2	2	3	3	3
W11_750	2	1	2	3	1	1	1	2	2	1	3	E37M51_365	3	3	3	3	3	3	3	3	3	3	3
W11_700	2	3	3	3	3	3	1	3	3	2	3	E37M51_320	3	3	2	3	3	3	3	3	3	3	3
W11_650	3	2	3	3	3	3	3	3	2	2	3	E37M51_317	3	3	3	2	3	3	3	3	3	3	3
W11_550	2	1	2	1	2	1	2	1	1	2	2	E37M51_297	3	3	3	3	2	3	3	2	3	3	3
W11_510	2	2	2	2	3	3	3	2	2	1	3	E37M51_295	3	3	3	3	3	3	3	3	3	3	3
K9_2600	2	3	1	2	1	1	2	1	2	2	2	E37M51_272	3	3	3	3	3	3	2	3	3	3	3
K9_1630	3	3	3	3	3	3	2	3	3	2	3	E37M51_255	1	3	2	2	3	2	3	3	2	2	3
K9_1050	3	2	3	3	2	2	2	3	3	2	3	E37M51_253	3	2	2	3	2	3	2	2	3	3	1
K9_1018	2	1	1	2	2	3	3	3	3	3	1	E37M51_215	2	2	2	1	1	3	2	3	1	2	1
K9_700	3	3	3	3	3	3	3	3	3	2	2	E37M51_207	2	3	2	2	2	3	3	3	3	3	3
K9_480	2	2	2	3	1	1	2	2	2	1	2	E37M51_190	3	2	3	3	2	3	2	3	3	2	2
K9_150	3	3	3	1	2	2	2	1	2	3	3	E37M51_188	3	3	2	3	3	3	3	3	2	3	3
X15_1630	2	3	3	3	3	3	3	2	3	2	3	E37M51_182	3	3	3	2	3	3	3	3	3	3	3
X15_950	3	3	2	3	2	1	3	2	3	2	3	E37M51_173	3	2	2	3	3	2	3	3	2	3	3
X15_750	2	3	3	3	3	3	2	3	2	3	3	E37M51_171	2	2	2	1	3	3	2	2	2	2	1
X15_600	1	3	3	3	3	2	3	2	3	3	2	E37M51_160	3	3	3	3	3	3	3	3	3	3	3
Y20_1100	3	2	3	2	3	3	3	3	3	3	3	E37M51_150	1	3	3	2	3	2	3	2	2	1	3
Y20_800	3	2	2	2	3	3	3	3	3	1	3	E37M51_125	2	2	2	2	2	2	1	2	3	2	3
Y20_550	2	2	2	2	2	3	2	3	3	3	2	E37M51_124	3	3	3	3	3	3	3	2	3	3	1
Y20_400	3	3	3	3	2	3	3	3	3	2	2	E37M51_117	3	3	3	3	3	3	3	3	3	3	3
												E37M51_99	2	3	3	2	3	3	3	3	2	3	3
												E37M51_90	1	2	2	1	2	2	2	3	3	2	3
												E37M51_89	3	3	3	3	3	2	3	3	3	3	3

Table 4 Estimates of multilocus outcrossing rates (t_m) single-locus outcrossing rate (t_s) and fixation index (F); a Student's *t* test gives the difference between estimates obtained with RADP and AFLP data

Item	t_m	t_s	t_m–t_s	F
RAPD	0.933 ± 0.027	0.923 ± 0.026	0.010	0.031 ± 0.000
AFLP	0.890 ± 0.033	0.828 ± 0.032	0.062	0.006 ± 0.000
t statistics	3.36	7.72		
P value	< 0.005	< 0.001		

populations of *E. urophylla* (House and Bell 1994). Our results further support the notion that the outcrossing rate of *E. urophylla* in exotic conditions is very similar to that in natural populations. They also indicate that estimates obtained with dominant marker data closely agree with those obtained from more traditional co-dominant isozyme data. In the case of dominant markers, however, instead of 3–5 polymorphic co-dominant isozymes, a larger number of marker loci is necessary to obtain a robust estimate (see below).

A Student *t* statistics was used to test whether a significant difference existed between the estimates of t_m obtained from RAPD and AFLP data. A similar approach was used by Shaw and Allard (1982) to study differences between upper-crown and lower-crown outcrossing rates in Douglas-fir. A significant difference was found both for t_m and t_s ($0.005 > P > 0.001$). The estimates obtained with RAPD markers were slightly higher than those obtained with AFLP. No a priori reason exists to justify such a difference; however, some hypotheses can be formulated to explain these results: (1) AFLP and RAPD have a different genetic base and sample different genomic regions, AFLP markers are based on the use of restriction enzymes and RAPD markers are based on amplified arbitrary sequences; (2) the difference could be specific to this experiment and due essentially to sampling error as different numbers of markers were used with AFLP and RAPD; (3) the difference could be due to the existence of missing data (1.02%) in the RAPD data set, while no missing data were seen in the AFLP data set. Although a significant difference in the estimates of t_m was found, the difference was only approximately 4%. In practical terms such a numerical difference does not change the main conclusion that outcrossing is highly predominant in the population.

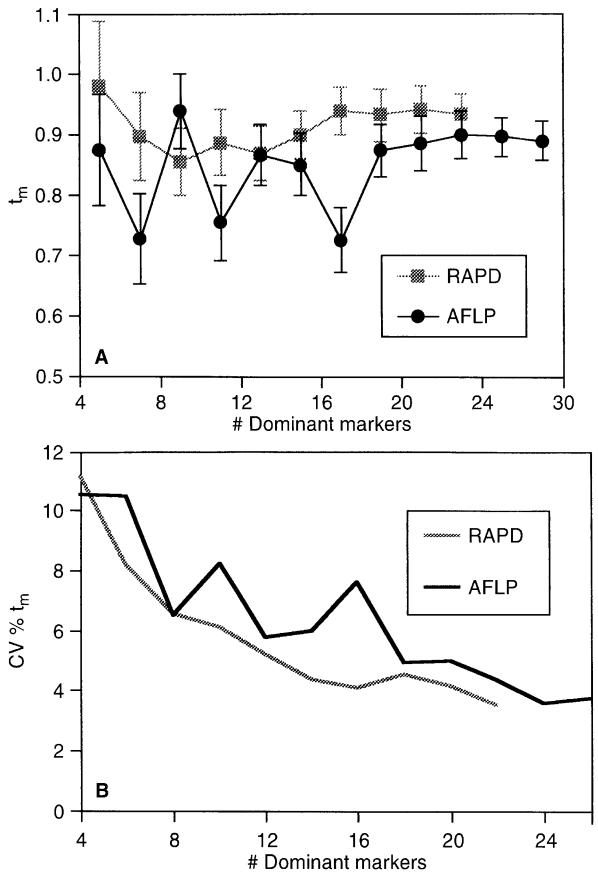


Fig. 3 Panel **A** Behavior of the estimate of outcrossing rate (t_m) and its respective standard error as a function of an increasing number of dominant markers, both RAPD and AFLP. Panel **B** Reduction in the coefficient of variation in the estimation of outcrossing rate as a function of the increasing number of dominant markers, both RAPD and AFLP

The fixation index, F, in the progeny estimated from AFLP data was lower than expected based on the estimate of t_m. For RAPD data no significant difference was found. Taking $t_m = 0.89$, the expected fixation index was $[F = (1 - t)/(1 + t)] = 0.05282$, while the estimated F was 0.006. A lower than expected F suggests an excess of heterozygotes and less inbreeding than expected in the progeny population analyzed. Contrary to our results, mating-system studies of natural populations of *Eucalyptus*, reported an F higher than expected based on the estimated t_m (Yeh et al. 1983; Coates and Sokolowski 1989; Peters et al. 1990; House and Bell 1994), indicating more inbreeding than expected in the progeny population used to carry out the study. These studies involved the use of isozyme markers assayed on embryos or cotyledons of ripe seeds before, or right at, germination. On the other hand, our study was carried out using rapidly growing 3-month-old plants grown in a greenhouse where low-vigor inbreeding-depressed seedlings either did not germinate or germinated poorly. It is very likely therefore that some level of selection against homozygous seedlings was unintentionally carried out when sampling plants for this study.

A χ^2 test indicated that observed progeny genotype frequencies did not conform to those expected under mixed mating for some marker loci. Several factors can contribute to such violations: selection against homozygous genotypes, genotype-dependent outcrossing rate, and the unbalanced frequencies of pollen in the population (Ritland 1983). In our study we found six RAPD markers and five AFLP markers showing significant deviations from the mixed-mating model (Table 2). These results support the hypothesis that selection against homozygous plants could have occurred in the nursery when sampling plants for the study, so resulting in a deficiency of homozygous genotypes.

Ritland and Jain (1981) in Fig. 6 of their original paper describing the mixed-mating model, show, through simulation studies, that between five and an infinite number of dominant marker loci with $P = 0.5$ are necessary to reach as low a variance on the estimate of outcrossing rate as that obtained with two co-dominant triallelic marker loci with $P = 0.33$. In our empirical analysis we observed that, with less than ten dominant markers with varying frequencies for the dominant allele, the estimates of t_m had two to three times the coefficient of variation compared to the estimates obtained with 18–20 markers (Fig. 3). These results therefore confirm previous simulation studies and build upon these by providing an empirical approximation and a recommendation as to the minimal number of dominant markers to be used for mating-system studies. This information will be of great utility in cases where the species under study does not display sufficient isozyme polymorphism, or does not have optimized isozyme-detection protocols such as in several tropical forest trees. Because RAPD and more recently AFLP markers have become more accessible providing large amounts of genetic data quickly with a universal set of reagents, we can anticipate an increasing use of such technologies in mating-system studies, particularly for those where isozyme systems have not been optimized.

Open-pollinated family breeding and outcrossing rate

The t_m estimated in this study gives a snap-shot estimate of the outcrossing events that took place in the original population of *E. urophylla*, a progeny trial, in the season prior to seed collection. The relevance of this estimate was in showing that the pattern of outcrossing versus selfing for *E. urophylla* in an exotic condition in Brazil did not differ significantly from that seen in the center of origin of the species in adjacent islands of Australia. From the *Eucalyptus* breeding stand point, where open-pollinated advancement of generations is typically the method of choice, it would also be very

interesting to make quantitative inferences as to the specific paternal composition of the open-pollinated offspring, and particularly on selected progeny individuals when carried out to the next generation of breeding. Such information could significantly help in deciding which, and how many, progeny individuals from the same O.P. family could be advanced. In this respect, future studies should be carried out with this breeding population, where phenological data, (establishing the coincidence of flowering periods and peaks amongst trees that compose the population), coupled to paternity exclusion studies with co-dominant multiallelic markers such as microsatellites (Brondani et al. 1997), could help quantifying this issue properly. It is reasonable to suggest, however, that a high outcrossing rate at the population level most likely reflects the participation of a large number of different pollen parents in the generation of outcrossed offspring, although full-sib relationships among trees within O.P. families might also exist.

It is important for the breeder to know not only how much outcrossing is taking place in a breeding population, but also how much selfing occurs. Based on the proportion of selfing, adjustments can be made to the coefficient of additive genetic variance that is estimated from open-pollinated families when the assumption of complete half-sib relationship is not met. In studies carried out with *Eucalyptus regnans*, Griffin and Cotterill (1988), suggested that with an outcrossing rate of 0.57 to 0.76, a coefficient equal to 1/2.5 should be used, instead of 1/4, to control the inflation of this parameter. Askew and El-Kassaby (1993) working with *Pinus taeda*, plotted the relationship between the % overestimation of the additive genetic variance and the outcrossing rate. Using this relationship for our *E. urophylla* population (10% selfing) a coefficient of 1/4 would cause an over-estimation of 15% in the additive genetic variance. Therefore, for the *E. urophylla* population studied, a coefficient of 1/3 should be used when estimating additive genetic variance, so as to avoid over-estimation of heritability and genetic gain.

Acknowledgments We acknowledge the support given by the Brazilian Ministry of Science and Technology through a PADCT-FINEP Competitve Grants in Biotechnology to D.G., the Brazilian National Research Council CNPq through a pre-M.Sc. Fellowship to F.A.G. and a Research Fellowship to D.G. Special thanks go to Ana Y. Ciampi, Cristina Marques and all the members of the Forest Biotechnology Group at North Carolina State University for their collaboration in the optimization of AFLP for *Eucalyptus*; additional thanks go to Márcio Gomes Squillassi and the personnel at Aracruz Celulose S.A for their help with plant collection.

References

Adams WT (1983) Applications of isozymes in tree breeding. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Part A. Elsevier Science Publishers, Amsterdam, pp 60–64

Askew GR, El-Kassaby YA (1993) Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. Theor Appl Genet 88:267–272

Brondani RPV, Brondani C, Tarchini R, Tingey S, Rafalski A, Grattapaglia D (1997) Discovery and development of microsatellite markers in *Eucalyptus*. Plant and Animal Genome V. San Diego California, abstract P276

Coates DJ, Sokolowski RE (1989) Geographic patterns of genetic diversity in Karri (*Eucalyptus diversicolor* F. Muell.). Aust J Bot 37:145–156

Fritsch P, Rieseberg LH (1992) High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. Nature 359:633–636

Gaiotto FA, Ciampi AY, Grattapaglia D (1997) AFLP in *Eucalyptus*: fast construction of linkage maps and estimation of mating system and genetic variation in a breeding population. Plant and Animal Genome V. San Diego, California, abstract P281

Grattapaglia D, Bradshaw H Jr(1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Can J For Res 24:1074–1078

Grattapaglia D, Sederoff RR (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudotestcross: mapping strategy and RAPD markers. Genetics 137:1121–1137.

Griffin AR, Cotteril PP (1988) Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. Silvae Genet 37:124–131

Hodge GR, Volker PW, Potts BM, Owen JV (1996) A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two eucalypt species. Theor Appl Genet 92:53–63

House APN, Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* S.T. Blake. Silvae Genet 43:167–176

Moran GF, Bell JC (1983) *Eucalyptus*. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Part B. Elsevier Science Publishers, Amsterdam pp 423–441

Namkoong G, Kang HC, Brouard JS (1988) Tree breeding: principles and strategies. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo

Peters BG, Lonie JS, Moran GF (1990) The breeding system, genetic diversity and pollen sterility in *Eucalyptus pulverulenta*, a rare species with small disjunct populations. Aust J Bot 38:59–70

Ritland K (1983) Estimation of mating systems. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Part A. Elsevier Science Publishers, Amsterdam, pp 289–302

Ritland K (1990) A series of FORTRAN computer programs for estimating plant mating systems. J Hered 81:235–237

Ritland K, Jain SA (1981) Model for the estimation of outcrossing rate and gene frequencies using "n" independent loci. Heredity 47:35–52

Shaw DV, Allard RW (1982) Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor Appl Genet 62:113–120

Vos P, Hogers R, Bleeker M, Reijans M, Lee TV, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acid Res 18:6531–6535

Yeh FC, Brune A, Cheliak WM, Chipman DC (1983) Mating system of *Eucalyptus citriodora* in a seed-production area. Can J For Res 13:1051–1055

Zobel BJ, Van Wyk G, Stahl P (1987) Growing exotic forests. John Wiley and Sons, New York